Data analysis software tools used during VIRGO engineering runs, review and future need

D. Buskulic

LAPP-EXP 2002-13
December 2002

Data analysis software tools used during VIRGO engineering runs, review and future need

D. Buskulic
On behalf of the Virgo Collaboration

LAPP-IN2P3-CNRS
9 chemin de Bellevue - BP. 110
F-74941 Annecy-le-Vieux Cedex

Presented at the VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2002), Moscow (Russia) June 24-28, 2002

Data analysis software tools used during Virgo engineering runs, review and future needs

D. Buskulic[a], on behalf of the VIRGO collaboration[b] *

[a]Laboratoire d'Annecy-le-Vieux de Physique de Particules,
B.P. 110, 74941 Annecy-le-Vieux CEDEX, FRANCE

[b]VIRGO Collaboration, http://www.virgo.infn.it

During last years, data flow and data storage needs for large gravitational waves interferometric detectors have reached an order of magnitude similar to high energy physics experiments. Software tools have been developed to handle and analyze those large amounts of data, with the specificities associated to gravitational waves search. We will make a review of the experience acquired during engineering runs on the VIRGO detector with the currently used data analysis software tools, pointing out the peculiarities inherent to our type of experiments. We will also show what the possible future needs for the Virgo data offline analysis are.

1. THE VIRGO EXPERIMENT

VIRGO [1][2] is an interferometric gravitational wave detection experiment. It consists mainly in a Michelson laser interferometer made of two orthogonal arms each of them being 3 kilometers long. Each arm consists of a Fabry-Perot cavity, extending the effective optical length up to 100 kilometers.

The experiment is a French-Italian collaboration of 11 laboratories and is located in Cascina, near Pisa in Italy. The commissioning of the full interferometer will start at the beginning of the year 2003.

1.1. Sensitivity

The Virgo sensitivity in gravitational wave strain h is shown on fig. 1 as a function of frequency. The frequency range in sensitivity extends from 4 Hz to a few kHz. The best sensitivity of about 3.10^{-23} $\mathrm{Hz}^{-1/2}$ should be attained around 500 Hz. Unlike data produced in high energy physics experiments, gravitational wave detection experiments produce a continuous stream of data. In the VIRGO case, the output signals are sampled at 20 kHz.

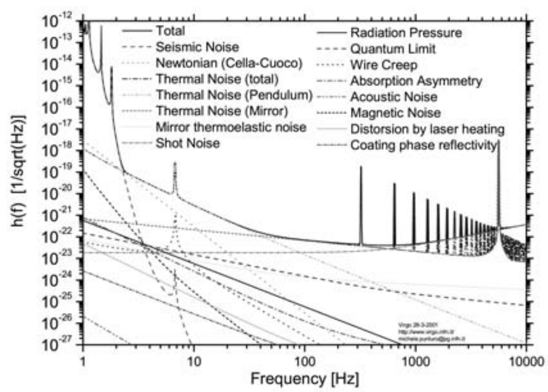

Figure 1. *The VIRGO sensitivity curve*

1.2. Noise reduction

In order to achieve the foreseen sensitivity, one has to fight all the noise sources, especially the seismic, thermal and optical noise. The main optical components, as well as injection and detection benches are suspended to mechanical super-attenuators that provide a seismic noise reduction of a factor 10^{12} at least, depending on frequency. Optical noise is reduced by the use of an ultra stable laser and mode cleaning cavities, at the input

*LAPP Annecy, LAL Orsay, ILGA Nice, ESPCI Paris, IPN Lyon, INFN Firenze, INFN Frascati, INFN Napoli, INFN Perugia, INFN Pisa, INFN Roma and at the output of the interferometer.

The VIRGO optical scheme is described elsewhere [1].

1.3. Data format

Most gravitational waves experiments in the world agreed on a common data format, the frame format [3]. The need for a common format arises from the need to exchange data among various experiments, as will be explained later. As has been pointed out, gravitational wave data is continuous. The frame format slices this data into frames representing a given time slice of interferometer data. In VIRGO, the frame length is 1 s. Thus, one frame contains ADC channels, monitoring data, reconstructed strain h, etc.

The full frame size coming out of the interferometer is about 4 MB. It is possible to build reduced frames containing only a subset of all channels or keep specified channels only for selected frames, thus making a preselection. The frame size is then reduced to approximately 100 kB.

1.4. Data acquisition system

The data acquisition system collects frame pieces from systems of the interferometer and builds final output frames. Control signals, monitoring signals, interferometer output, reconstructed signals and trigger signals are collected.

All the data are then sent to various data displays and stored to disk. The DAQ system gives a raw data rate of 4 MB/s, or 9 MB/s uncompressed. Taking into account dead time, this will give 100 to 125 TB/year of accumulated raw data.

The online processing system part of the DAQ produces digested (trend data) or down sampled (50 Hz) data.

1.5. Engineering runs in 2001-2002

The central part of the interferometer, consisting of all the systems except the two end mirrors of the two 3 km arms form a small, 6 m Michelson interferometer. This central interferometer (CITF) was tested and commissioned during the construction and completion of the two long arms, from February 2001 to July 2002. This allowed to make control tests, various systems tests and to build necessary tools for the future.

Five engineering runs were performed on the CITF during this period. They lasted 3 days each, allowing the collection of 3 TB of data each time. A lot was learnt about the machine. Control procedures were tested and used, stability of operation was monitored, data collection and the whole data acquisition system were debugged. Data display tools have been developed and the first analyzes on real data were made.

After engineering runs, the raw data were sent to computing centers through the network.

The data analysis software tools mainly used were a home made Data Display, the well known analysis environment Matlab and a home made analysis environment based on ROOT [5] named VEGA [6]. A few uses of the PAW environment were also made.

1.6. Offline computing

The size of the foreseen data set imposes a distributed computing model. Two computing centers, the CCIN2P3 in Lyon and the computing center in Bologna are foreseen to analyze VIRGO data. The computing resources at the site, in Cascina, will also be used. A central bookkeeping database will be used to manage the data.

2. DATA ANALYSIS CHALLENGES

In order to understand the specificities of our data analysis, it is interesting to describe some of the data analysis challenges we are faced with.

2.1. Pulsar searches

Rotating neutron stars that have a small asymmetry and spin around an off-symmetric axis generate a gravitational wave. This wave gives a very weak signal, of amplitude $h \lesssim 10^{-25}$ which is buried in the noise. However, this signal is steady and one needs to integrate it over a very long period, of the order of a few months, to extract the signal from noise.

The signal is quasi-periodic, the variation in frequency is due to the Earth movement and the Earth-star relative movement. The characteristics of the signal are dominated by the star parameters, like rotation period, quadrupolar moment. The signal also depends on the position of the star in the sky[7]. For one given set of parameters (star position, period,...) one can do a search in the interferometer output. But taking into account the whole parameter space, one would end up with 10^{29} different cells[7] meaning 10^{16} TFlops of needed computing power. Fortunately, this number may be decreased to 1 TFlops by using hierarchical methods.

2.2. Coalescing binaries searches

Another analysis challenge is the search for coalescing binary compact objects. A binary system of massive objects emits gravitational waves. In the case of compact objects, like neutron stars, this emission induces a fall down inspiralling phase, followed by a merger (coalescence) that is detectable by gravitational waves detectors. The inspiral and merger last from a few seconds to a minute within the VIRGO sensitivity bandwidth. The gravitational wave strain is of the order of $h \approx 10^{-23}$ to 10^{-22}, which is still small compared to the foreseen sensitivities. The expected signal may be calculated, the shape and values of the signal depend mainly on the masses of the two stars.

The search for a known signal buried in noise is done using the optimal filtering technique (or Wiener filtering) [8]. One calculates a signal for a given set of parameters, in our case the two star masses, then performs a weighted intercorrelation (optimal filter) between the output signal of the interferometer and the calculated signal. The search has to be done for all possible physical parameters of the system, giving theoretical signals called templates. The number of templates necessary for a full search in the parameter space of interest was estimated to be around 5.10^5 imposing a needed processing power of 300 GFlops.

2.3. Distributed computing

It is possible in both analyzes to distribute computing. In the case of pulsar searches with hierarchical methods, each node analyzes one frequency band or one subset of the parameter space. For coalescing binaries, each node deals with a subset of all templates, which is equivalent to a subset of the parameter space.

3. SOFTWARE DATA ANALYSIS TOOLS

We will present the main software data analysis tools used during VIRGO engineering runs.

3.1. Data Display

The Data Display is an online and offline monitoring and display tool. It allows access to files or remote frames and includes a channel browser (fig. 2). It is able to read and display frame files content, as well as receive and display frame contents sent over the network (fig.3). The graphical user interface was built using the Xforms library[9], while the display of data uses the ROOT[5] libraries.

Figure 2. *The channel browser in the Data Display*

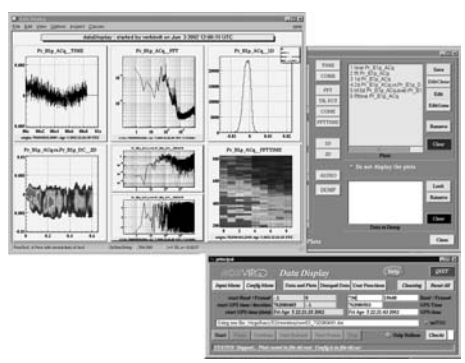

Figure 3. *The graphical output of a few channels in the Data Display*

3.2. Matlab

The commercial Matlab tool [10] is very popular among the gravitational wave search community, due to its rich set of signal analysis functions and tools. It was interfaced with frame files, allowing the knowledge and experience gathered with this environment to be used extensively. However, the amount and complexity of data collected and foreseen is such that there is a need for another environment, sometimes used in conjunction with Matlab, which would allow easy management of large amounts of data.

3.3. VEGA

VEGA (Visual Environment for Gravitational waves data Analysis) [6] is such an environment. It is based on the ROOT environment and addresses the data analysis, data handling and signal processing needs of the community. The scripting is performed through the built-in CINT interpreter, while the ability to visualize time dependent data needed an adaptation of the ROOT visualization classes. The possibility to display time series up to a million points was verified. VEGA includes signal processing capabilities. This is done through interfaces to external libraries, such as FFTW [11] and a home made signal processing library.

Access to data may be done through direct calls to the Frame handling library, FrameLib[4]. For more complex cases, data usually consists of the frame files and of meta information containing the location, summary, trigger information about the files. A bookkeeping database may be considered as meta information database. In VEGA, the data is accessed through a "frame channel" which acts as a data provider. It first looks at meta information (location, trigger) to further access the requested data. Meta information may be a central bookkeeping database or a local metadatabase built by VEGA. This allows the handling of a local set of files.

3.4. Example : trend data visualization

An example of the use of the tools described above may be found in the visualization of trend data on the web. Trend data are summary data down sampled at 1 Hz which are very useful to see the status of the interferometer over a long time period. In order to see this status locally or remotely, web pages are generated in quasi-realtime showing a few relevant time plots.

The trend frame data files are recorded on disk and a local metadatabase is updated by a simple VEGA script. Plots are generated every 30 minutes by a set of VEGA scripts which are controlled by a global shell script.

The relevant plots concern the general status of the interferometer (e.g. power output), alignment signals, suspensions damping signals, detection bench alignment and output, laser and first (input) mode cleaner signals, and finally data acquisition statistics.

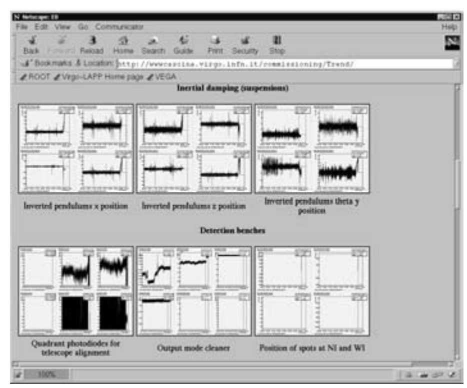

Figure 4. *Web page showing some trend data*

The resulting plots are accessed through a web page (fig.4) by people on shift or remotely by people of the collaboration who are interested. This showed to be very flexible, yet simple, and useful for long term visual monitoring of the interferometer status. Thanks to the existing and previously developed tools, like VEGA data access and management and plotting capabilities, the development of trend data visualization was quite fast.

4. COOPERATIVE ANALYSIS AND DATA EXCHANGE

When a gravitational wave comes to earth, it may a priori be seen by all detectors. Whether it is seen or not depends mainly on its orientation and amplitude. A cooperative analysis among several detectors scattered on the earth surface allows to extract more information from the signal, such as the direction of propagation or some physical parameters of the object that generated the wave.

This leads to the need to exchange data between experiments. The common frame format was designed in this context. Some online monitoring data were already exchanged in quasi-realtime between the two experiments VIRGO and LIGO[12] for testing purposes.

For the future there is a wish to use the GRID tools for data exchange, after the remaining problems in middleware compatibility between the European DataGrid (for VIRGO) and the American iVDGL (for LIGO) are solved.

In VIRGO, some tests were made to evaluate the suitability of GRID tools for the kind of analyzes of interest, in the European DataGrid context.

A test of binary coalescence search, where each node deals with a subset of the parameter space, and a test of pulsar search, where each node analyzes a frequency band were made[13]. It was verified that multiple jobs can be submitted and the output retrieved with small overhead time. The conclusion is that computational grids seem suitable to perform data analysis for coalescing binaries and periodic sources searches.

5. CONCLUSIONS

Gravitational wave detection experiments produce a large amount of data, in the 100 TB per year range. This data is of continuous nature and a common format was designed for it, the frame format. The analysis of this data needs sizeable computing power, in the range 300 GFlops to 1 TFlops.

The data analysis tools used during VIRGO engineering runs were presented, Data Display, Matlab and VEGA, as well as an example of use, the trend data visualization on the web.

There is a need for exchanging data among gravitational wave detection experiments, which was already tested at the level of monitoring data between the two experiments VIRGO and LIGO. Finally, there are some efforts under way to test and use the coming GRID tools.

REFERENCES

1. VIRGO Coll., Final Design Report, 1997
2. See http://www.virgo.infn.it
3. Joint LIGO/VIRGO note LIGO-T970130-B and VIRGO-SPE-LAP-5400-102. See also B.Mours, Gravitational Wave Detection, Proc. TAMA International Workshop on Gravitational Wave Detection, Saitama, Japan, November 12-14 1996, K. Tsubono, M.K. Fujimoto, K Kuroda, eds. (1997) 27
4. See http://wwwlapp.in2p3.fr/virgo/FrameL
5. *ROOT - An Object Oriented Data Analysis Framework*
   Rene Brun and Fons Rademakers, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/.
6. D. Buskulic et al.; International Journal of Modern Physics D, Vol. 9, No 3 (2000) 293-297. See also http://wwwlapp.in2p3.fr/virgo/vega
7. *Periodic sources*, S. Frasca, Virgo-SIGRAV School on Gravitational Waves 2002, See also http://www.sigrav.unige.it/scuole/cascina/ /2002/default.htm
8. See, e.g., N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications* (Wiley, New York, 1949) or Vainshtein, L. A. and Zubakov, V. D. *Extraction of Signals from Noise* (New York, Dover, 1970)
9. See http://world.std.com/ xforms/
10. See http://www.mathworks.com/ /products/matlab/
11. See http://www.fftw.org
12. *Network data analysis server (NDAS) prototype*, S.Márka, B. Mours and R. Williams, Class. Quantum Grav. **19**(2002) 1537-1540
13. *A Grid Approach to Geographically Distributed Data Analysis for Virgo*, C. Palomba, G. Tortone et al, GWDAW 2002 Workshop